ID	Suggested Revision	Revisions/Notes
1-1	The English is good but there are still some minor issues here and there. Please ask a native English speaker to proofread your paper. For example, the abstract contains several grammatical mistakes.	Text and grammar changes throughout (see tracked changes).
1-2	"This article describes the design and evaluation of a novel educational programming game, Pirate Plunder, to teach procedural abstraction to children age 10 and above." In several places in the paper readers can see different age groups (e.g., 5 to 11, 10 and above, etc.). Try to be more consistent and keep in mind that you may not easily be able to claim your game is effective for adults as well. That being said, I suggest authors to replace "above" with a more appropriate number perhaps.	Removed "age 10 and above" from section 1.
1-3	"Despite these skills being an important part of computer science, children, particularly in primary education (age 5 to 11), are rarely taught them because of a lack of teacher expertise, curriculum materials and questions over whether they are cognitively developed enough to understand them." Such claims should be back up by appropriate references.	Removed cognitive development claim and added two references to support the others in that sentence (section 1).
1-4	"The article builds upon a previous report on the design of Pirate Plunder (anon for blind review) by describing in more detail how the game introduces the learning content." In the reference, authors should have mentioned the venue that this work is under review but had not.	Removed the word "provisionally" which could suggest the work is not yet published and is under review (section 1). To clarify for the reviewers, the referenced work has been published.
1-5	"There are a multitude of educational programming tools available for use in computer science education." I recommend authors to cite some of these tools here: for example:	Added citations for three examples of educational programming tools (including the example given in the suggested revision) (section 1.1.1).

	- Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. Computers & Education, 94, 18-36.	
1-6	"This type of block-based programming is also prevalent in other popular tools used in primary education including Code.org, Tynker, Hopscotch and Purple Mash." Please include corresponding references.	Added web page references for each of these products (section 1.1.1), all of which are commercially produced.
1-7	Page 5, section 1.4: I expect authors to talk about existing educational games for computer programming and possibly computational thinking that are in line with their study. Authors should talk about their strength and weaknesses and then try to build up upon these existing games in the literature, rather than ignoring the existing literature and gaps in the area. For instance, I name an adaptive game that has recently been developed for promoting computational thinking. The game uses different terms for promotion of abstraction (pattern recognition). I think it would be fair to name some of their tools before calling your game a novel game AutoThinking: An Adaptive Computational Thinking Game	Added a section of 'Related work', part of which explains how procedural abstraction is introduced existing games (section 1.4.1). Including a screenshot of procedures in Lightbot. Then stated how Pirate Plunder aims to build on these, feeding into the key features of the game design.
1-8	Research questions, that drive the paper, should be built in the introduction from an ongoing and pertinent bibliography (up to 2020). These should be of global interest and not focused to a particular local problem. Identifying a research gap is not enough; key is showing its significance to the field.	Moved research questions to the start of the paper and mentioned them in the relevant parts of the introduction. Added 'Related work' (section 1.4.1), part of which discusses similar efforts and clarifies the research gap and significance to the field.
1-9	Section 1.4.2: In Fig 2 what we can see is a short hint,	Moved the explanation of tutorials, feedback, and hints into separate paragraphs.

	it would be interesting if authors actually differentiate hints, feedback, and tutorials. Very confusing in its current form. I also recommend authors to replace Fig 8 with another Fig to better show readers how the feedback is given, where the hint is, how instruction (tutorials) are given and etc.	Changed the caption on Figure 8 (now Figure 9) to explain where the tutorial instructions and feedback are given. Added extra figure (10) that shows hints being given on a challenge level and a table (2) that explains the hints given on the first procedures challenge.
1-10	"These were based on player questions and researcher answers recorded during game testing." You'd better add respective reference here or delete this.	Removed sentence (section 1.4.3).
1-11	"Based on the literature review, the research questions are as follows:" How can you claim that these RQs are driven from literature. As I mentioned before, your literature review is neither systematic, nor holistic (its more arbitrary and several credible works in the area have been left out). Please either fix an appropriate literature review section or rewrite this sentence	Changed sentence to "The article aims to answer the following research questions:" (now moved to section 1).
1-12	In the methodology indicate, clearly describe, and justify with references: participants; case study; sample and its representativeness; the instruments used and its validation; how the data you are looking for match your objectives; your experimental process; the statistical methods.	Split the 'Materials' section into 'Intervention materials' with sub-sections for each condition (section 2.3) and 'Instruments and measures' with sub-sections for each pre/post-test measure (section 2.4).

Changed 'Hypotheses' section to 'Data analysis' (section 2.6) that now includes the hypotheses (section 2.6.1) and the measures that relate to each, the statistical methods used to test these (section 2.6.2), and the measures used address the research questions (section 2.6.3). |
| 1-13 | Page 14: before conducting the one-way ANCOVA, authors should report homogeneity of regression to show whether or not the knowledge level of groups were similar/comparable. Please do the same in section 3.2.2 and other sections | Added homogeneity of regression slopes to section 3.1.1.

As this is the only ANCOVA (the rest of the tests are one-way ANOVAs or independent samples t-tests) homogeneity of regression has not been added to the other sections. |
| 1-14 | I also advise authors to update Table 5 and add all necessary statistics in it | Not sure what the reviewer means by this. The table contains the descriptive statistics of the mean, N and standard deviation for each group. Which additional statistics are |

		required?
1-15	Section 4.1.2, I assume authors are aware of the fact that CT is not programming (coding) and is about conceptualization. I understand that they try to use programming-based environments as a vehicle to deliver CT, but as several credible studies (including those reported by Wing the founder of CT) have clearly stated that environment using coding syntax should not be used to deliver CT, I suggest authors to add a few lines of discussion and state such limitation (you may want to use the following references for this) -Understanding Computational Thinking before Programming: Developing Guidelines for the Design of Games to Learn Introductory Programming through Game-Play. - AutoThinking: An Adaptive Computational Thinking Game	Added a section (1.1.2) to the introduction about computational thinking, acknowledging its wider context and stating that this paper will concentrate on the use of these skills in computer programming. Added extra detail on the Computational Thinking test and stated that ideally this would be combined with a measure that does not use programming syntax (section 2.4.4). Added sentence in computational thinking results section to state limitations with using a programming-based measure (section 4.2).
1-16	"If the game is too difficult, then the player will become anxious and if it is too easy, then they will become bored. Pirate" is this flow state? If yes, you may want to update your sentence	Removed suggested sentence and replaced it with other flow enablers (clear goals and immediate and accurate feedback). Also edited proceeding sentence as given data does not actually show that players were in this flow state (section 4.3.1).
1-17	"As such, we recommend that educational games that are designed to be used in the classroom are evaluated against the resources that they would be replacing or used alongside." There exists several research studies reporting similar experimental studies (educational games vs traditional technology-enhanced learning approaches). What do you mean by "improving evaluations of educational games"? I recommend authors to name a few of these studies first and then state that there exist a lot of studies	Rewritten section to mention good examples of studies first, before going on to give the common weaknesses of designs and make the recommendation (section 4.4.3).

	ignoring such designs (it seems better than neglecting effort of other researchers in the field)	
2-1	A methodology section is needed that involves the following subsections: research design, participants, measurements, procedure, data analysis, and results sections. The methodology needs to make clear what you are going to measure and how. Please consider to remove information that do not really give anything to the paper.	See revision for 1-12.
2-2	The introduction needs to identify the research area, to mention similar efforts, and to underline the reason of performing this research (research gap). Then, it is very important to explain from the beginning how you are going to perform your research (e.g. we used three groups, two control and one experimental in order to measure ...).	See revision for 1-8.
2-3	Finally, your paper needs a related work section that builds on your research questions (I suggest a paragraph for each RQ) and a discussion section that discusses your results, answering these research questions.	Added RQ to the title of sections 4.1 and 4.3 that address them and moved computational thinking (now section 4.2) into its own section as it is not part of the research questions.
2-4	There are three groups in a between groups experiment; one group that attended non programming curricula (?), one group that played the game, and one group that attended scratch programming lessons. I think that spreadsheets do not involve scratch programming curricula. I am not also sure if the students of the scratch programming group were taught clones and custom blocks(?) like the game group. A between groups experiment needs the 3 groups to be taught the same curricula with different ways / tools. If I missed it, please provide a table to describe the curricula of the 3 approaches.	Added an explanation of why those control groups were used (section 2.3.3) and why they were not taught custom blocks and clones. Added a sentence to identify this as future work (section 5.1).

	Otherwise, I am not sure what the between groups experiment examines?	
2-5	In the second phase, the researchers asked the 2 control groups to also play the game. Authors describe each groups results throughout the different approaches. If the game's approach was the only one that teaches clones and custom blocks, that maybe explains why the students were improved using that. I understand that the authors wanted to give all the students the opportunity to play the game but they need to explain why this information must be included in the paper. Please make things more clear to the reader.	Added an additional explanation of why those control groups were used and why they were not taught custom blocks and clones (section 2.3.3). Explanation of why all three groups played the game (section 2.2) has been expanded to include generating game analytics from more players to address RQ2.
2-6	On the contrary, the game's evaluation includes interesting information that need a better description in the different sections of the paper. Methodology does not make clear the measuring instruments of "why the game was effective". I suggest that the authors must emphasize on these aspects and remove unnecessary things.	Added section 2.4.6 to explain the game analytics data and how this is used to address why the game was effective (RQ2).
2-7	The use of a game based approach in order to teach difficult programming curriculum could be interesting for computing education. Perhaps a better description of how the game succeeds on this is needed. Maybe in comparison with similar games, if there are any.	See revision for 1-7.
2-8	Finally, the game's design may exist in another research but it would be useful to include some of this information too.	Added more detail about the game design (section 1.4) and how it differs from other programming games (section 1.4.1). If reviewers have any other specific areas of the game design in mind that they thin would be useful to include, then it would be good to know what these are.